PERGAMON Computers and Mathematics with Applications 41 (2001) 1033–1041

An International Journal
computers & mathematics with applications

www.elsevier.nl/locate/camwa

Fitting a Least Squares Piecewise Linear Continuous Curve in Two Dimensions

S. KUNDU
Computer Science Department
Louisiana State University
Baton Rouge, LA 70803, U.S.A.
kundu@bit.csc.lsu.edu

V. A. UBHAYA
Department of Computer Science and Operations Research
258 IACC Building
North Dakota State University
Fargo, ND 58105, U.S.A.
vasant_ubhaya@ndsu.nodak.edu

(Received November 1999; accepted January 2000)

Abstract—An optimal piecewise linear continuous fit to a given set of n data points $D = \{(x_i, y_i) : 1 \leq i \leq n\}$ in two dimensions consists of a continuous curve defined by k linear segments $\{L_1, L_2, \ldots, L_k\}$ which minimizes a weighted least squares error function with weight w_i at (x_i, y_i), where $k \geq 1$ is a given integer. A key difficulty here is the fact that the linear segment L_j, which approximates a subset of consecutive data points $D_j \subset D$ in an optimal solution, is not necessarily an optimal fit in itself for the points D_j. We solve the problem for the special case $k = 2$ by showing that an optimal solution essentially consists of two least squares linear regression lines in which the weight w_j of some data point (x_j, y_j) is split into the weights λw_j and $(1-\lambda)w_j$, $0 \leq \lambda \leq 1$, for computations of these lines. This gives an algorithm of worst-case complexity $O(n)$ for finding an optimal solution for the case $k = 2$.

Keywords—Least squares regression, Nonlinear regression, Piecewise linear continuous curve, Convexity, Optimization, Algorithms, Complexity.

1. INTRODUCTION

Any continuous curve can be approximated arbitrarily closely by a piecewise linear continuous (in short, multilinear) curve. Hence, the problem of fitting a multilinear least squares curve to a given set of data points $D = \{p_i : 1 \leq i \leq n\}$, where $p_i = (x_i, y_i)$ are points in the plane, is useful when we do not have *a priori* knowledge of the shape of the curve to be fitted. An instance of this problem arises in the construction of fuzzy rules of the form "if x is in A_j, then y is c_j", where A_j is a fuzzy set with a triangular membership function and c_j is a constant [1,2]. See

0898-1221/01/$ - see front matter Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
PII: S0898-1221(00)00337-0

also [3,4]. The problem is also of independent mathematical interest requiring applications of optimization theory.

The k-Linear Fit Problem

Assume $x_1 < x_2 < \cdots < x_n$, where $p_i = (x_i, y_i)$. For a given integer $k \geq 1$, let $\{L_1, L_2, \ldots, L_k\}$ denote an approximating multilinear curve where the successive line segments $L_1, L_2, \ldots, L_k$ of the curve are listed from left to right. We allow for the possibility that the two adjacent segments L_j and L_{j+1} have the same slope, and thus, they are parts of the same line. Let q_{j-1} and q_j be the left and right end points of L_j, where $q_j = (u_j, v_j)$, $0 \leq j \leq k$, with $u_0 = x_1$ and $u_k = x_n$. Let $D_j = \{p_i : x_i \in [u_{j-1}, u_j]\}$, $1 \leq j \leq k$. Clearly, $\cup\{D_j : 1 \leq j \leq k\} = D$. Also $D_j \cap D_{j+1} = \{p_i\}$ for some i if L_j and L_{j+1} intersect on a vertical line through p_i or equivalently $u_j = x_i$. Otherwise, $D_j \cap D_{j+1} = \emptyset$. Thus, $\{D_1, D_2, \ldots, D_k\}$ is a decomposition of D, where two adjacent sets may contain at most one common point. We call D_j the support set of L_j. We assume that each D_j is nonempty. Now, we let $y_i' = L_j(x_i)$ if $p_i \in D_j$. Note that y_i' are uniquely defined, for if $p_i \in D_j \cap D_{j+1}$, then $u_j = x_i$ and $L_j(u_j) = L_{j+1}(u_j)$ since $q_j = (u_j, v_j)$ is a point on both L_j and L_{j+1}. Our goal is to find a multilinear curve $\{L_1, L_2, \ldots, L_k\}$ so as to minimize the error

$$E = \sum \left\{ w_i \left(y_i - y_i'\right)^2 : 1 \leq i \leq n \right\},$$

where $w_i > 0$ is the weight associated with p_i, $1 \leq i \leq n$. Loosely speaking, the line segment L_j approximates points in D_j. Later on, we will see that if $D_j \cap D_{j+1} = \{p_i\}$, then optimality considerations may require us to include the point p_i in both L_j and L_{j+1} with different weights.

Since each D_j is nonempty, by assumption, we must have $k \leq n$. The problem is trivial for $k = n$ and $n - 1$. For $k = n$, we choose a line segment L_j through each point p_j so that segments through adjacent points intersect strictly between them. Clearly, $D_j = \{p_j\}$ and $E = 0$. For $k = n - 1$, we do a similar construction except that the last segment passes through points p_{n-1} and p_n. Then, $D_j = \{p_j\}$, $1 \leq j \leq n - 2$, $D_{n-1} = \{p_{n-1}, p_n\}$, and $E = 0$. Alternately, we may let L_j be the line segment joining p_j and p_{j+1}, $1 \leq j \leq n - 1$. Then, $D_j = \{p_j, p_{j+1}\}$, $1 \leq j \leq n - 1$, and $E = 0$. Henceforth, we will assume that $k < n - 1$.

Given a multilinear curve, we may combine its contiguous line segments of the same slope into a single line segment obtaining another multilinear curve with fewer line segments. Conversely, if the support set D_j of a line segment L_j of a multilinear curve has two or more elements, then we may break L_j into smaller line segments with their support sets having at least one element.

In this article, we analyze the two-linear fit problem (Sections 2 and 3), and develop an algorithm of worst-case complexity $O(n)$ to compute an optimal fit (Section 4). One of the basic observations here is that in an optimal solution $\{L_1, L_2\}$ with each L_i having a support set D_i, an L_i may not be an optimal fit for the corresponding points in D_i. In fact, if $D_1 \cap D_2 = \{p_j\}$ for some j, then we show that an optimal solution $\{L_1, L_2\}$ indeed consists of two regression lines, one on D_1 and another on D_2 with the weight w_j of p_j split into λw_j for D_1 and $(1 - \lambda)w_j$ for D_2 for some $0 \leq \lambda \leq 1$. We obtain an efficient procedure for computing such a λ. Since n points p_i are accessed for computing an optimal fit, our $O(n)$ algorithm is clearly optimal to a multiplicative constant.

A multidimensional generalization of the above problem arises when each x_j is an m-dimensional point, i.e., $x_j \in R^m$ $(m > 1)$, and we wish to fit a polyhedral surface to D. For $m = 2$, this means that we have to partition the plane R^2 into a set of k polygons (one polygon for each subset of points D_j) and define a plane L_j on that polygon in such a way that these planes form a continuous surface; i.e., they intersect each other on the vertical planes along the boundaries of the polygons such that the total error E is minimized. The general problem remains unsolved at this point and may require a different technique for its solution. Several other curve fitting problems of mathematical and practical interest appear in one of the author's earlier works in [5–7], and other references given therein.

2. BASIC OBSERVATIONS AND RESULTS

First, we briefly review the convexity properties of the classical regression problem [8]. Given $p_i = (x_i, y_i)$, $1 \le i \le n$, we minimize

$$E(s,t) = \sum \{w_i[y_i - (s + tx_i)]^2 : 1 \le i \le n\}$$

to find the optimal values of s and t in the regression line $y = s + tx$. Since x_i are distinct and $w_i > 0$, it is easy to show using second partial derivatives that E is a strictly convex function of (s,t) for $n \ge 2$ [9]. Consequently, the unique optimal values of s and t are obtained by letting $\frac{\partial E(s,t)}{\partial s} = \frac{\partial E(s,t)}{\partial t} = 0$. They are as follows:

$$s = \frac{\left(\sum w_i x_i^2\right)\left(\sum w_i y_i\right) - \left(\sum w_i x_i\right)\left(\sum w_i x_i y_i\right)}{\left(\sum w_i\right)\left(\sum w_i x_i^2\right) - \left(\sum w_i x_i\right)^2}, \tag{2.1}$$

$$t = \frac{\left(\sum w_i\right)\left(\sum w_i x_i y_i\right) - \left(\sum w_i x_i\right)\left(\sum w_i y_i\right)}{\left(\sum w_i\right)\left(\sum w_i x_i^2\right) - \left(\sum w_i x_i\right)^2}, \tag{2.2}$$

where the summation $\sum$ is over the set $\{1 \le i \le n\}$. We use these formulae in Section 4. If $n = 1$, then any line passing through $p_1 = (x_1, y_1)$ is optimal. It may be easily seen that the optimal s and t can be computed by an algorithm of worst-case complexity $O(n)$, where n is the number of points p_i.

We now state the following known result as a lemma for future reference.

LEMMA 2.1. *Let L_1 and L_2 be the two least squares regression lines for the sets of points $D_1 = \{p_i : 1 \le i \le j\}$ and $D_2 = \{p_i : j+1 \le i \le n\}$, respectively. Suppose that the two lines are parallel, meet, and form a single line; then the single line is the least squares regression line for $D = D_1 \cup D_2$.* ∎

THEOREM 2.2. *Let the lines $\{L_1, L_2\}$ constitute an optimal two-linear ($k = 2$) fit. Assume that the lines are not parallel and meet at a point $q = (x', y')$ strictly between vertical lines through p_j and p_{j+1}, i.e., $x_j < x' < x_{j+1}$, or they are parallel, meet, and form a single line. In either case, then the support sets $D_1 = \{p_i : 1 \le i \le j\}$ and $D_2 = \{p_i : j+1 \le i \le n\}$. Then, L_1 and L_2 are the least squares regression lines for D_1 and D_2, respectively. (In case $j = 1$, L_1 is the line joining q to p_1, and similarly for L_2 when $j + 1 = n$.) If the two lines are parallel and form a single line, then this line is the least squares regression line for $D = D_1 \cup D_2$.*

PROOF. First, consider the case when the two lines are not parallel. Then, by the definition of support sets, D_i are as stated. Suppose that $L_1 : y = s + tx$ is not the least squares regression line for D_1. Let

$$E_1(s,t) = \sum \{w_i[y_i - (s + tx_i)]^2 : 1 \le i \le j\}$$

be the error corresponding to L_1. First, suppose that $j \ge 2$; i.e., D_1 contains at least two points. Since L_1 is not the regression line, we must have $\frac{\partial E_1(s,t)}{\partial s} \ne 0$ or $\frac{\partial E_1(s,t)}{\partial t} \ne 0$. Hence, there exists some point (s', t') such that $(\frac{\partial E_1(s,t)}{\partial s})(s' - s) + (\frac{\partial E_2(s,t)}{\partial t})(t' - t) < 0$. It follows that we may choose a point (s'', t'') on the line joining (s,t) and (s', t') in the plane so that the following holds: (s'', t'') is sufficiently close to (s,t), the x-coordinate v of the point of intersection of the line $y = s'' + t''x$ (call it L_1'') with L_2 satisfies $x_j < v < x_{j+1}$, and, by convexity or otherwise, $E_1(s'', t'') < E_1(s,t)$. This shows that the lines $\{L_1'', L_2\}$ have a smaller total error than those of $\{L_1, L_2\}$ contradicting the optimality of $\{L_1, L_2\}$. Thus, L_1 is the regression line for D_1.

Now assume that $j = 1$. Then, $D_1 = \{p_1\}$. In this case, the regression line for D_1 will pass through p_1 and will have zero error. Since L_1 is not the regression line by hypothesis, we have $E_1 > 0$. It follows that, if L_1 is replaced by a suitable line L_1' passing through p_1 and intersecting L_2 appropriately, then $\{L_1', L_2\}$ will provide a better approximation. This contradicts the optimality of $\{L_1, L_2\}$.

Now consider the case when the two lines L_1 and L_2 are parallel and form one line. Then they may be assumed to meet at any point strictly between vertical lines through p_j and p_{j+1}. Then, by the definition of support sets, D_i are as stated. Again, by the above proof, they are the regression lines for D_1 and D_2, respectively.

The last statement of the theorem follows from Lemma 2.1. ∎

3. MAIN RESULTS

In Theorem 2.1, we considered the case when the intersection point q of the two lines L_1 and L_2 of an optimal two-linear fit lies strictly between vertical lines through points p_j and p_{j+1}. In Theorem 3.1 below, we consider the case when q lies on the vertical line through a point p_j. The two regression lines in this case are obtained by splitting the weight w_j of p_j into λw_j and $(1-\lambda)w_j$ as will be shown below. To this effect, define the errors of two lines $y = s_1 + t_1 x$ and $y = s_2 + t_2 x$ by

$$E_1(s_1, t_1, \lambda) = \sum_{1 \le i \le j-1} w_i [y_i - (s_1 + t_1 x_i)]^2 + \lambda w_j \left[y_j - (s_1 + t_1 x_j)\right]^2, \tag{3.1}$$

$$E_2(s_2, t_2, \lambda) = (1-\lambda) w_j \left[y_j - (s_2 + t_2 x_j)\right]^2 + \sum_{j+1 \le i \le n} w_i [y_i - (s_2 + t_2 x_i)]^2, \tag{3.2}$$

where $0 \le \lambda \le 1$. We now state our theorem.

THEOREM 3.1. *Suppose that the lines $L_1 : y = \alpha_1 + \beta_1 x$ and $L_2 : y = \alpha_2 + \beta_2 x$ form an optimal two-linear fit such that the following holds. The lines are not parallel (i.e., $\beta_1 \ne \beta_2$), and they intersect at a point $q = (x_j, y_j')$ on the vertical line through p_j where $y_j' = \alpha_1 + \beta_1 x_j = \alpha_2 + \beta_2 x_j$, so that the support sets $D_1 = \{p_i : 1 \le i \le j\}$ and $D_2 = \{p_i : j \le i \le n\}$. Then there exists $0 \le \lambda \le 1$ such that L_1 is the regression line for the points in D_1 with the weight w_j of $p_j = (x_j, y_j)$ modified to λw_j, and L_2 is the regression line for the points in D_2 with the weight w_j of p_j modified to $(1-\lambda)w_j$ with the weights of all other points $\{p_i : i \ne j\}$ remaining unchanged at w_i. Moreover, λ is unique if $y_j - (\alpha_1 + \beta_1 x_j) = y_j - (\alpha_2 + \beta_2 x_j) \ne 0$, i.e., $y_j' \ne y_j$, and λ can have any value in $[0,1]$ if $y_j' = y_j$.*

PROOF. Let $g_1(s_1, t_1) = s_1 + t_1 x_j$ and $g_2(s_2, t_2) = s_2 + t_2 x_j$. Also, let $E_1(s_1, t_1, \lambda)$ and $E_2(s_2, t_2, \lambda)$ be as defined by (3.1) and (3.2). Then it is easy to see that, for *each* fixed λ, $0 \le \lambda \le 1$, $E_1(\alpha_1, \beta_1, \lambda) + E_2(\alpha_2, \beta_2, \lambda)$ gives the total error due to $\{L_1, L_2\}$, and hence, it is the minimum value of the objective function

$$E_1(s_1, t_1, \lambda) + E_2(s_2, t_2, \lambda), \tag{3.3}$$

subject to $g_1(s_1, t_1) - g_2(s_2, t_2) = 0$. Since each $E_i(s_i, t_i, \lambda)$ is convex and each $g_i(s_i, t_i)$ is linear in (s_i, t_i), the following Karush-Kuhn-Tucker optimality conditions are both necessary and sufficient [9,10]. For each $0 \le \lambda \le 1$, there exists a constant $u = u(\lambda)$ such that the following equations hold at $(s_1, t_1, s_2, t_2) = (\alpha_1, \beta_1, \alpha_2, \beta_2)$:

$$\frac{\partial E_1}{\partial s_1} + u(\lambda) \frac{\partial g_1}{\partial s_1} = \frac{\partial E_1}{\partial s_1} + u(\lambda) = 0, \tag{3.4}$$

$$\frac{\partial E_1}{\partial t_1} + u(\lambda) \frac{\partial g_1}{\partial t_1} = \frac{\partial E_1}{\partial t_1} + u(\lambda) x_j = 0, \tag{3.5}$$

$$\frac{\partial E_2}{\partial s_2} - u(\lambda) \frac{\partial g_2}{\partial s_2} = \frac{\partial E_2}{\partial s_2} - u(\lambda) = 0, \tag{3.6}$$

$$\frac{\partial E_2}{\partial t_2} - u(\lambda) \frac{\partial g_2}{\partial t_2} = \frac{\partial E_2}{\partial t_2} - u(\lambda) x_j = 0. \tag{3.7}$$

Here, both (3.4) and (3.6) show that $u(\lambda)$ is a linear function with $u(\lambda) = A + 2B\lambda$, where $B = w_j[y_j - (\alpha_1 + \beta_1 x_j)] = w_j[y_j - (\alpha_2 + \beta_2 x_j)]$. Suppose first that there exists a $\lambda \in [0,1]$ such that $u(\lambda) = 0$; clearly, λ is unique if $B \neq 0$. Equations (3.4)–(3.7) then become $\frac{\partial E_i}{\partial s_i} = 0, \frac{\partial E_i}{\partial t_i} = 0$, $i = 1, 2$, and hence, both L_1 and L_2 are the regression lines as desired in the theorem.

Now, if there is no λ for which $u(\lambda) = 0$, assume that $u(\lambda) > 0$ for all $\lambda \in [0,1]$. By linearity of $u(\lambda)$, this is equivalent to both $u(0)$ and $u(1) > 0$. We show that this leads to a contradiction. The case $u(\lambda) < 0$ for all $\lambda \in [0,1]$ is handled in a similar way. First, suppose that $\beta_1 > \beta_2$. Then, for all sufficiently small $\delta > 0$, the line $L_1(\delta) : y = (\alpha_1 + \delta) + \beta_1 x$, which is parallel to L_1, will intersect the line L_2 slightly to the left of the previous intersection point q of L_1 and L_2, which is on the vertical line through p_j. We let $\lambda = 0$. Then, in the objective function in (3.3), the point (x_j, y_j) is included in the line L_2 with its full weight w_j as it should be. However, since $u(0) > 0$, (3.4) with $\lambda = 0$ gives us $\frac{\partial E_1}{\partial s_1} < 0$. This shows that the lines $\{L_1(\delta), L_2\}$ will give a lower value of the objective function in (3.3) than that given by $\{L_1, L_2\}$. This contradicts the optimality of $\{L_1, L_2\}$. Now, suppose that $\beta_1 < \beta_2$. Then, for all sufficiently small $\delta > 0$, the line $L_2(-\delta) : y = (\alpha_2 - \delta) + \beta_2 x$ will intersect L_1 slightly to the right of the point q. We let $\lambda = 1$ so that the point (x_j, y_j) is correctly included in L_1 in the objective function in (3.3) with its full weight w_j. However, with $\lambda = 1$ in (3.6) with $u(1) > 0$, we get $\frac{\partial E_2}{\partial s_2} > 0$. Hence, the lines $\{L_1, L_2(-\delta)\}$ give a smaller value of the objective function than that given by $\{L_1, L_2\}$. This is again a contradiction. This shows that the linear function $u(\lambda)$ must vanish for some λ in $[0,1]$.

If $B = 0$, i.e., $y_j = \alpha_1 + \beta_1 x_j = \alpha_2 + \beta_2 x_j = y'_j$, then it is a simple property of regression that for any λ in $[0,1]$, the regression lines L_1 and L_2 will remain unchanged and will meet at (x_j, y_j). (In this case, $u(\lambda) = 0$ gives $A = 0$, and hence, $u(\lambda)$ identically equals 0.) ∎

One may wonder why equations (3.5) and (3.7) do not play a direct role in the above argument. This is because we can assume without loss of generality that $x_j = 0$, by shifting the coordinate system along the x-axis. This does not change the geometry of the points and the lines $\{L_1, L_2\}$, nor does it change the values of the coefficients β_1 and β_2.

In light of the above theorem and Theorem 2.2, we now formulate a restricted version of the two-linear fit problem where the two lines, parallel or otherwise, meet at some point $q = (x', y')$ such that $x_j \leq x' \leq x_{j+1}$ for some fixed j. Recall that $E_1(s_1, t_1, \lambda)$ and $E_2(s_2, t_2, \lambda)$ are defined by (3.1) and (3.2), respectively.

SUBPROBLEM SP_j, $1 \leq j \leq n-1$ (Restricted Two-Linear Fit Problem). Find s_1, t_1, s_2, t_2 (i.e., lines $y = s_1 + t_1 x$ and $y = s_2 + t_2 x$), point (x', y'), and λ so as to

$$\begin{aligned}
&\text{minimize } E_1(s_1, t_1, \lambda) + E_2(s_2, t_2, \lambda),\\
&\text{subject to } y' = s_1 + t_1 x' = s_2 + t_2 x',\\
&\qquad x_j \leq x' \leq x_{j+1}, \quad 0 < \lambda \leq 1, \quad (1-\lambda)(x' - x_j) = 0.
\end{aligned}$$ ∎

The case $x_j < x' < x_{j+1}$ in the subproblem SP_j corresponds to Theorem 2.2. In this case, by the definition of the support sets, we have $D_1 = \{p_i : 1 \leq i \leq j\}$ and $D_2 = \{p_i : j+1 \leq i \leq n\}$. Since $x_j < x'$, the last constraint of SP_j shows that $\lambda = 1$. Hence, p_j in D_1 is assigned full weight w_j as it should be. On the other hand, the case $x' = x_j$ in SP_j corresponds to Theorem 3.1. In this case, by the definition of the support sets, $D_1 = \{p_i : 1 \leq i \leq j\}$ and $D_2 = \{p_i : j \leq i \leq n\}$. We have $0 < \lambda \leq 1$, and p_j is assigned weights λw_j and $(1-\lambda)w_j$ in D_1 and D_2, respectively. (When $\lambda = 1$, these weights are w_j and 0.) The case $\lambda = 0$ is not included in SP_j; it is covered by SP_{j-1}. In this case, the regression lines for $\{p_i : 1 \leq i \leq j-1\}$ and $\{p_i : j \leq i \leq n\}$ meet at x_j. The following proposition is obvious.

PROPOSITION 3.2. *If $\{L_1, L_2\}$ is an optimal two-linear fit, then $\{L_1, L_2\}$ is also optimal to subproblem SP_j for some $1 \leq j \leq n-1$.* ∎

Our next theorem, which concerns the optimal solutions of the subproblem SP_j, will aid us in the development of an algorithm to solve the two-linear fit problem.

THEOREM 3.3. *Let L_1 and L_2 be the two least squares regression lines for the set of points $D_1 = \{p_i : 1 \le i \le j\}$ and $D_2 = \{p_i : j+1 \le i \le n\}$, respectively. If the two lines are not parallel and meet at (u,v) with $x_j \le u \le x_{j+1}$, or are parallel, meet, and form a single line, then the lines give an optimal solution to the subproblem* SP_j. *In the latter case, the single line is the regression line for $D = D_1 \cup D_2$.*

PROOF. First, suppose that L_1 and L_2 are not parallel. Clearly, L_1, L_2 with $(x', y') = (u, v)$ and $\lambda = 1$ (which ensures that x_j in D_1 carries full weight w_j) form a feasible solution to SP_j. We show that they are optimal to SP_j. Let Δ_1 and Δ_2 be the errors of L_1 and L_2, respectively. Suppose that the lines $y = s_1 + t_1 x$, $y = s_2 + t_2 x$, with (x', y') as in the statement of SP_j, be any feasible solution to SP_j. Suppose first that $x_j < x' < x_{j+1}$. Then, by the last constraint of SP_j, we see that $\lambda = 1$. Since Δ_1 and Δ_2 are minimal errors, we have

$$E_1(s_1, t_1, 1) = \sum_{1 \le i \le j} w_i [y_i - (s_1 + t_1 x_i)]^2 \ge \Delta_1,$$

$$E_2(s_2, t_2, 1) = \sum_{j+1 \le i \le n} w_i [y_i - (s_2 + t_2 x_i)]^2 \ge \Delta_2.$$

Hence,

$$E_1(s_1, t_1, 1) + E_2(s_2, t_2, 1) \ge \Delta_1 + \Delta_2,$$

which shows the optimality of L_1 and L_2. Now, suppose that $x' = x_j$. Then, λ may take any value in $(0, 1]$. In this case, by a constraint on SP_j, we have $s_1 + t_1 x_j = s_2 + t_2 x_j$. Using this fact, we obtain as before,

$$E_1(s_1, t_1, \lambda) + (1-\lambda) w_j \left[y_j - (s_2 + t_2 x_j)\right]^2 = \sum_{1 \le i \le j} w_i [y_i - (s_1 + t_1 x_i)]^2 \ge \Delta_1,$$

$$E_2(s_2, t_2, \lambda) - (1-\lambda) w_j \left[y_j - (s_2 + t_2 x_j)\right]^2 = \sum_{j+1 \le i \le n} w_i [y_i - (s_2 + t_2 x_i)]^2 \ge \Delta_2.$$

Thus, $E_1 + E_2 \ge \Delta_1 + \Delta_2$, showing the optimality of L_1 and L_2. Now, consider the case that L_1 and L_2 are parallel and form one line. Then the lines may be assumed to meet at some (u, v) with $x_j \le u \le x_{j+1}$. Then the proof as above applies, and we conclude that L_1 and L_2 are optimal for SP_j. The last statement of the theorem follows from Lemma 2.1. ∎

4. ALGORITHM AND COMPUTATIONS

In this section, we develop an algorithm (Algorithm 4.2) of worst-case complexity $O(n)$ for obtaining an optimal solution of the two-linear fit problem. We first obtain several formulae for computation which will lead us to an $O(n)$ algorithm.

The regression lines minimizing $E_1(s_1, t_1, \lambda)$ and $E_2(s_2, t_2, \lambda)$ defined by (3.1) and (3.2) obviously depend upon j and λ. We denote them by $y = s_{1,j}(\lambda) + t_{1,j}(\lambda)x$ and $y = s_{2,j}(\lambda) + t_{2,j}(\lambda)x$, respectively. Next, we compute $s_{i,j}(\lambda)$ and $t_{i,j}(\lambda)$, $1 \le i \le 2$. To this effect, define constants for the left regression line for $2 \le j \le n-1$, by

$$a_{1,j} = \sum_{\{1 \le i < j\}} w_i, \qquad b_{1,j} = \sum_{\{1 \le i < j\}} w_i x_i, \qquad c_{1,j} = \sum_{\{1 \le i < j\}} w_i x_i^2,$$

$$d_{1,j} = \sum_{\{1 \le i < j\}} w_i, y_i, \qquad h_{1,j} = \sum_{\{1 \le i < j\}} w_i x_i y_i, \qquad k_{1,j} = \sum_{\{1 \le i < j\}} w_i y_i^2.$$

Now, applying (2.1) and (2.2) appropriately, rearranging the terms, and *suppressing* the index j from our notations for convenience, we obtain

$$s_{1,j}(\lambda) = s_1(\lambda) = \frac{(c_1 d_1 - b_1 h_1) + \left(-h_1 x + d_1 x^2 + c_1 y - b_1 x y\right)\lambda w}{(a_1 c_1 - b_1^2) + (c_1 - 2b_1 x + a_1 x^2)\lambda w},$$

$$t_{1,j}(\lambda) = t_1(\lambda) = \frac{(a_1 h_1 - b_1 d_1) + (h_1 - d_1 x - b_1 y + a_1 x y)\lambda w}{(a_1 c_1 - b_1^2) + (c_1 - 2b_1 x + a_1 x^2)\lambda w}.$$

It is easy to show that $a_{1,j}c_{1,j} - (b_{1,j})^2 > 0$ and $c_{1,j} - 2b_{1,j}x_j + a_{1,j}x_j^2 > 0$ for $3 \le j$. Similarly, we define constants $a_{2,j} = \sum_{\{j<i\le n\}} w_i$, $b_{2,j} = \sum_{\{j<i\le n\}} w_i x_i$, etc., for the right regression line for $2 \le j \le n-1$ symmetrically as above. Then we obtain $s_{2,j}(\lambda) = s_2(\lambda)$ and $t_{2,j}(\lambda) = t_2(\lambda)$ from the above expressions for $s_{1,j}(\lambda)$ and $t_{1,j}(\lambda)$ by substituting a_2, b_2, c_2, d_2, and h_2 for a_1, b_1, c_1, d_1, and h_1, respectively, and *replacing* λ with $1-\lambda$. By expanding $E_1(s_1, t_1, \lambda)$ and $E_2(s_2, t_2, \lambda)$, it is easy to show that

$$E_1(s_1,t_1,\lambda) = E_1 = k_1 + a_1 s_1^2 + c_1 t_1^2 - 2d_1 s_1 - 2h_1 t_1 + 2b_1 s_1 t_1 + \lambda w_j \left[y_j - (s_1 + t_1 x_j)\right]^2,$$
$$E_2(s_2,t_2,\lambda) = E_2 = k_2 + a_2 s_2^2 + c_2 t_2^2 - 2d_2 s_2 - 2h_2 t_2 + 2b_2 s_2 t_2 + (1-\lambda) w_j \left[y_j - (s_2 + t_2 x_j)\right]^2.$$

Again, E_1 and E_2 depend upon j, but we suppress j for simplicity.

When $\lambda = 1$, the lines $y = s_{1,j}(1) + t_{1,j}(1)x$ and $y = s_{2,j}(1) + t_{2,j}(1)x$ give the regression lines for the sets $D_1 = \{p_i : 1 \le i \le j\}$ and $D_2 = \{p_i : j+1 \le i \le n\}$, respectively, since the point p_j is included in D_1 with full weight w_j. When $\lambda = 1$, $s_{1,j}(\lambda)$ and $t_{1,j}(\lambda)$ may be easily simplified to the following:

$$s_{1,j}(1) = s_1(1) = \frac{c_{1,j+1}d_{1,j+1} - b_{1,j+1}h_{1,j+1}}{a_{1,j+1}c_{1,j+1} - b_{1,j+1}^2},$$

$$t_{1,j}(1) = t_1(1) = \frac{a_{1,j+1}h_{1,j+1} - b_{1,j+1}d_{1,j+1}}{a_{1,j+1}c_{1,j+1} - b_{1,j+1}^2}.$$

Similarly,

$$s_{2,j}(1) = s_2(1) = \frac{c_{2,j}d_{2,j} - b_{2,j}h_{2,j}}{a_{2,j}c_{2,j} - b_{2,j}^2},$$

$$t_{2,j}(1) = t_2(1) = \frac{a_{2,j}h_{2,j} - b_{2,j}d_{2,j}}{a_{2,j}c_{2,j} - b_{2,j}^2}.$$

Note that indices $j+1$ and j, respectively, appear on the right sides of these two sets of numbers. Again, when $\lambda = 1$, we may obtain

$$\begin{aligned} E_1(s_1,t_1,1) = E_1 &= k_{1,j+1} + a_{1,j+1}s_{1,j+1}^2 + c_{1,j+1}t_{1,j+1}^2 - 2d_{1,j+1}s_{1,j+1} - 2h_{1,j+1}t_{1,j+1} \\ &\quad + 2b_{1,j+1}s_{1,j+1}t_{1,j+1}, \\ E_2(s_2,t_2,1) = E_2 &= k_{2,j} + a_{2,j}s_{2,j}^2 + c_{2,j}t_{2,j}^2 - 2d_{2,j}s_{2,j} - 2h_{2,j}t_{2,j} + 2b_{2,j}s_{2,j}t_{2,j}. \end{aligned}$$

Note that $s_{1,j}(\lambda)$ and $t_{1,j}(\lambda)$ for λ in $[0,1]$ can be written in the form $A/(B + C\lambda) + D$, with $B > 0$ and $C > 0$, if $j \ge 3$, and $s_{2,j}(\lambda)$ and $t_{2,j}(\lambda)$ in the form $A/(B - C\lambda) + D$, with $B > C > 0$, if $n - j \ge 2$, where A, B, C, and D do not contain λ. In the latter, the coefficient of λ is -1 since λ in $s_{1,j}(\lambda)$ and $t_{1,j}(\lambda)$ is replaced by $1-\lambda$ to obtain $s_{2,j}(\lambda)$ and $t_{2,j}(\lambda)$. Note that the function $u(\lambda) = A/(B + C\lambda)$ with $B > 0$, $C > 0$ is

(i) strictly decreasing and strictly convex if $A > 0$, and
(ii) strictly increasing and strictly concave if $A < 0$.

Similarly, the function $v(\lambda) = A/(B - C\lambda)$ with $B > C > 0$ is

(i) strictly increasing and strictly convex if $A > 0$, and
(ii) strictly decreasing and strictly concave if $A < 0$.

The values of λ for which the two regression lines meet on a vertical line through x_j may be easily obtained by solving the equation $s_{1,j}(\lambda) + t_{1,j}(\lambda)x_j = s_{2,j}(\lambda) + t_{2,j}(\lambda)x_j$. Substituting for $s_{1,j}(\lambda)$ and $t_{1,j}(\lambda)$, and simplifying, we obtain, after *suppressing* the index j,

$$s_1(\lambda) + t_1(\lambda)x = \frac{(c_1 d_1 - b_1 h_1) + (a_1 h_1 - b_1 d_1)x - \left(a_1 c_1 - b_1^2\right) y}{(a_1 c_1 - b_1^2) + (c_1 - 2b_1 x + a_1 x^2)\,\lambda w} + y.$$

Similarly, $s_2(\lambda) + t_2(\lambda)x$ equals the right-hand side of the above equation with index 1 replaced by 2 and λ by $1 - \lambda$. Equating these two right-hand sides, we obtain

$$\frac{(c_1d_1 - b_1h_1) + (a_1h_1 - b_1d_1)x - (a_1c_1 - b_1^2)\,y}{(a_1c_1 - b_1^2) + (c_1 - 2b_1x + a_1x^2)\,\lambda w} = \frac{(c_2d_2 - b_2h_2) + (a_2h_2 - b_2d_2)x - (a_2c_2 - b_2^2)\,y}{(a_2c_2 - b_2^2) + (c_2 - 2b_2x + a_2x^2)\,(1-\lambda)w}, \tag{4.1}$$

which gives the values of λ for which the two regression lines meet on a vertical line through x_j. Note that (4.1) is a linear equation in λ. The following proposition may be easily established by using (4.1). The simple proof is left to the reader.

PROPOSITION 4.1. *Exactly one of the following applies to the two regression lines $y = s_{1,j}(\lambda) + t_{1,j}(\lambda)x$ and $y = s_{2,j}(\lambda) + t_{2,j}(\lambda)x$.*

(a) *For a unique λ in $[0, 1]$, the two lines meet at (x_j, y_j'), where $y_j' \neq y_j$.*
(b) *For every λ in $[0, 1]$, the two lines meet at (x_j, y_j).*
(c) *There exists no λ in $[0, 1]$ such that the two lines meet on a vertical line through x_j.* ∎

The above proposition confirms the observations made in Theorem 3.1.

We now state our algorithm.

ALGORITHM 4.2 (for computing an optimal two-linear fit).

STEP 1. Compute the regression line $L_2 : y = s_{2,1}(1) + t_{2,1}(1)x$ on the set $\{(x_i, y_i) : 2 \le i \le n\}$. Let $E = E_2(s_2, t_2, 1)$ be the associated sum of squared errors for L_2. Let L_1 be any line through (x_1, y_1) which meets L_2 at a point (x', y'), where $x_1 \le x' \le x_2$. This line has zero error. Call the pair $\{L_1, L_2\}$ the current optimal solution.

STEP 2. For $j = 2, 3, \ldots, n-1$ do the following.

(a) Compute the regression lines $L_1 : y = s_{1,j}(1) + t_{1,j}(1)x$ on $D_1 = \{(x_i, y_i) : 1 \le i \le j\}$ and $L_2 : y = s_{2,j}(1) + t_{2,j}(1)x$ on $D_2 = \{(x_i, y_i) : j+1 \le i \le n\}$. For $j = n-1$, let L_2 be any line through (x_n, y_n) which meets L_1 between x_{n-1} and x_n.
(b) If L_1 and L_2 are not parallel and meet at (x', y') where $x_j \le x' \le x_{j+1}$, or are parallel, meet, and form one single line, then compute the errors $E_1 = E_1(s_1, t_1, 1)$ and $E_2 = E_2(s_2, t_2, 1)$ for these lines. Let $E' = E_1 + E_2$. For $j = n-1$, let $E_2 = 0$ and $E' = E_1$, the error of L_1. (In the case of parallel lines, the single line is the regression line for $D = \{(x_i, y_i) : 1 \le i \le n\}$.) If $E' < E$, then replace E by E' and let the two regression lines $\{L_1, L_2\}$ be the current optimal solution.
(c) (In this case $x' < x_j$ or $x_{j+1} < x'$.) Compute λ from the linear equation (4.1). If λ is unique and satisfies $0 < \lambda < 1$, then compute the regression lines $L_1 : y = s_{1,j}(\lambda) + t_{1,j}(\lambda)x$ on $D_1 = \{(x_i, y_i) : 1 \le i \le j\}$ with the weight λw_j for (x_j, y_j), and $L_2 : y = s_{2,j}(\lambda) + t_{2,j}(\lambda)x$ on $D_2 = \{(x_i, y_i) : j \le i \le n\}$ with the weight $(1-\lambda)w_j$ for (x_j, y_j). The weights at all other points (x_i, y_i), $i \neq j$ equal w_i. Compute the errors $E_1 = E_1(s_1, t_1, \lambda)$ and $E_2 = E_2(s_2, t_2, \lambda)$, and let $E' = E_1 + E_2$. If $E' < E$, then replace E by E' and let the two regression lines $\{L_1, L_2\}$ be the current optimal solution. ∎

At termination, the above algorithm gives an optimal fit. The proof of the correctness of the algorithm follows from the results of Section 3. By Proposition 3.2, an optimal solution of the two-linear fit problem is also an optimal solution of the subproblem SP_j for some $1 \le j \le n-1$. The algorithm finds a solution, which minimizes the least squares objective among all optimal solutions of SP_j for all $1 \le j \le n-1$. Steps 1 and 2(a),(b) are based on Theorem 3.3. For each j, they find an optimal solution of SP_j corresponding to $\lambda = 1$. This solution, if it exists, is in the form of two regression lines for points $D_1 = \{p_i : 1 \le i \le j\}$ and $D_2 = \{p_i : j+1 \le i \le n\}$, which meet on a vertical line between p_j and p_{j+1}. If the two regression lines do not meet between p_j and p_{j+1}, then we use Step 2(c) to find that solution of SP_j, if it exists, for which $0 < \lambda < 1$, as justified by Theorem 3.1. Values of λ are obtained by solving (4.1); two different possibilities as stated in Proposition 4.1(a) and (b) apply to λ. Note that it is sufficient to consider $\lambda = 1$ to cover the entire case (b) of Proposition 4.1. Since $\lambda = 1$ is already considered in Steps 1 and 2(a),(b) of the algorithm, this leaves only a part of case (a) of Proposition 4.1 for which $0 < \lambda < 1$. This is considered in Step 2(c) of the algorithm. Thus, all values of λ are considered, and the algorithm terminates with an optimal solution.

Using the formulae already derived, it is straightforward to show that the algorithm is of worst-case complexity $O(n)$. We may compute $a_{1,j}$, for all $2 \le j \le n-1$, by one forward pass in $O(n)$ time using the following recursive relations: $a_{1,1} = 0$, $a_{1,j} = a_{1,j-1} + w_{j-1}$, $2 \le j \le n-1$. In the same way, we may compute $b_{1,j}$, $c_{1,j}$, $d_{1,j}$, $h_{1,j}$, $k_{1,j}$ and $a_{2,j}$, $b_{2,j}$, $c_{2,j}$, $d_{2,j}$, $h_{2,j}$, $k_{2,j}$, for all $2 \le j \le n-1$, in $O(n)$ time.

Now, we consider Steps 1 and 2(a),(b) of the algorithm. Using the already computed above numbers, we may determine $s_{i,j}(1)$, $t_{1,j}(1)$, and $E_i(s_i, t_i, 1)$, $1 \le i \le 2$, for a fixed j in $O(1)$ time by applying the special formulae derived above for $\lambda = 1$ for these quantities. In Step 2(c), we may compute λ, $s_{i,j}(\lambda)$, $t_{i,j}(\lambda)$, and $E_i(s_i, t_i, \lambda)$, $1 \le i \le 2$, for a fixed j in $O(1)$ time again using the formulae derived above for an arbitrary λ. Since Step 2 is repeated $n-2$ times, the algorithm is $O(n)$.

Finally, we remark that all the computations for the algorithm may be organized alternatively by considering regression lines of the form $y = s + t(x - x_j)$, for each j, rather than $y = s + tx$ as we have done. This leads to somewhat simpler formulae at the expense of more computations. Nevertheless, the complexity remains at $O(n)$.

REFERENCES

1. G.J. Klir and T.A. Folger, *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall, Englewood Cliffs, NJ, (1988).
2. C.-T. Lin and C.S.G. Lee, *Neural Fuzzy Systems*, Prentice Hall, Upper Saddle River, NJ, (1996).
3. C.C. Lee, Fuzzy logic in control systems: Fuzzy logic controller—Part II, *IEEE Trans. Syst. Man. Cybern.* **20**, 419–435, (1990).
4. X.-J. Zeng and M.G. Singh, Approximation theory of fuzzy systems—SISO case, *IEEE Trans. on Fuzzy Systems* **2**, 162–176, (1994).
5. M.-H. Liu and V.A. Ubhaya, Integer isotone optimization, *SIAM J. Optim.* **7**, 1152–1159, (1997).
6. V.A. Ubhaya, An $O(n)$ algorithm for least squares quasi-convex approximation, *Computers Math. Applic.* **14** (8), 583–590, (1987).
7. V.A. Ubhaya, Duality and Lipschitzian selections in best approximation from nonconvex cones, *J. Approx. Theory* **64**, 315–342, (1991).
8. J. Neter, M. Kutner, W. Wasserman and C. Nachtshein, *Applied Linear Statistical Models, Fourth Edition*, Irwin, Columbus, OH, (1996).
9. R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, (1970).
10. M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming, Theory and Algorithms*, John Wiley and Sons, New York, (1993).